小特集「知的メディア検索技術の動向」

知的映像メディア検索技術の動向

Intelligent Retrieval for Video Media

美濃 導彦*
Michihiko Minoh

* 京都大学工学部附属高度情報開発実験施設
Integrated Media Environment Experimental Laboratory, Kyoto University.

1995年9月18日 受理

Keywords: intelligent retrieval, video media, multimedia, pattern recognition.

1. はじめに

マルチメディアという言葉は，さまざまな意味で用いられるが，ここでは，音声，画像，文字データなどの異なるメディアを統合的に扱う技術の総称として用いる．

マルチメディアデータは，マルチメディアで扱われるデータの総称で，その中心は動画像と音声が時間的に同期している映像である．本稿では，静止画や動画像の一フレームを「画像」，静止画の時系列を「動画像」，動画像に同期した音声データを含めて「映像」という用語を用いる．

人間どうしのコミュニケーションは時間と空間を共有することが原点であった．すなわち，同じ時間に同じ空間にいることが出合いであり，コミュニケーションの始まりである．この制約のもとでは人間は五感をすべて活用できるので，マルチメディア処理によるコミュニケーションが可能である．

しかし，見方を変えればこの制約は大きな負荷となる．すなわち，空間的に近くにいる同時代の人間とコミュニケーションができないからである．

メディア変遷技術は，時間と空間の制約を取り除くことに寄与してきた．紙の発明に始まり，写真，テープレコーダ，ビデオレコーダ，計算機のディスクといった記録技術の進展が時間的な制約を取り除くことを可能にした．電気通信の技術とコンピュータネットワーク，ワークステーションの発達が空間的制約を取り除いた．その結果，インターネットという巨大なマルチメディア*1空間が世界的規模で構築されつつある*2．

音声，画像，映像などのマルチメディアデータは，基本的には，センサによる出力をディジタル化しただけであり，形式的には膨大な数字の配列になる．意味的な情報はそのなかに埋め込まれており，これを取り出す技術が，一般的にはパターン認識と呼ばれている．

認識技術をメディア処理に活用して新しい視点を開こうとする試みが，「知能情報メディア」研究会*3で行われているが，そのなかの重要課題の一つが映像の知的処理である[池山 94，田村 95]．

データベースの分野でも映像を対象にしたオブジェクト指向データベースやマルチメディアデータベースの議論が活発に行われている[Grosky 94]．

このように，映像を対象にした入力処理技術，認識技術，蓄積技術，検索技術，出力処理技術は今後ますます重要になってくる．

本稿では，認識技術を利用した知的映像処理および，認識結果を利用した知的検索技術について，現在の動向を概観する．

＊1 マルチメディアといっても，厳密には，嗅覚，味覚は抜けている．人間の入力情報の90%程度は視聴覚からの情報であるといわれているので，ここでは除外して考えるが，これらの感覚がなければ本当の意味でのマルチメディアではないかもしれない．

＊2 メディアのディジタル化技術が，コンピュータでこれらの情報を統一的に扱ううえでの基礎になっているのはいうまでもない．

＊3 電子情報通信学会第3種研究会（平成5年10月～平成7年9月），第2種研究会（予定：平成8年4月～平成10年3月）．

2. 映像検索の特徴

2・1 映像情報空間

映像は動画像と音声が同期したものである．空間的には，画像を2次元，音声を1次元，合わせて3次元で，これに時間軸が加わった4次元スペースである．これは，ディジタル化された映像を信号として捉えた考え方であり，計算機に蓄えられる膨大な数字の配列の次元数が4であるといっているだけである．

これにもう1次元，情報抽出の次元を考える．図1に示すように，これは，従来のパターン認識における情報の抽象化の軸である．一番低いレベルは信号のレベル（数字の羅列）で，次の段階が画像信号処理や音声信号処理により計算機が抽出できる特徴レベルの記述である．特徴レベルからさらに抽象化を進めようとするとセマンティックギャップにつきあたる．その向こうにあるのが人工知能の分野で扱われているシンボル記述のレベル（テキスト表現，言語表現）である．映像を見て人間がキーワードを与える場合は，人間がこの処理をするので，直接シンボルレベルの情報が得られる．計算機にこの処理を行わせる場合は，特徴レベルとシンボルレベルの橋渡しをする「モデル」が必要である．情報抽出の次元は，マルチメディアデータに，検索のためのインデックスを付加するものである．

映像空間は，図2に示すような5次元空間で表現できる．画像の2軸と，音声の1軸は，信号がそのまま

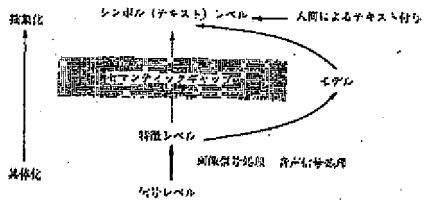

図1 情報抽出軸

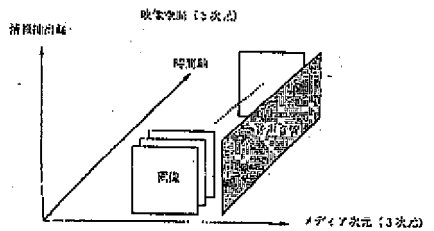

図2 映像空間

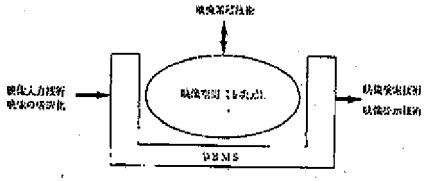

図3 映像データベース

人る軸であり，時間軸に対しては同期がとれている．現在のところ，画像は画像，音声は音声の信号として処理されている．一般的には，音声と画像を信号レベルで統合する処理は意味づけが困難であり，現在まで行われていない．しかし，将来的には試みる価値がある問題である．

時間軸を持つメディアは，その場で処理するのと，記録したものを処理するのとで，ほとんど同じ時間が必要である．例えば，講演記録を考えてもらいたい．これをテープレコーダに録音したものを真剣に聞こうとすると講演と同じ時間がかかる．これが，活字としてまとまった文書になると一覧性が高まり，飛ばし読みができるので，時間の制約から解放される．

情報抽出軸に対するデータは，もとのマルチメディアデータを処理して生成するか，人間がキーワードを付加することにより生成される．この軸のデータの質が，知的な検索を実現するうえでのキーポイントになる．

図3に示すように，映像空間に代表されるマルチメディア空間にデータを入力する処理，空間に存在するデータを蓄積する処理，蓄積されているデータを検索・提示する出力処理のすべてが，マルチメディアデータベース，特に，映像データベースには必要である．

本稿では，特に，入力側での映像の構造化，検索側での映像検索の手法，結果の提示方法に重点をおいて述べる．映像の蓄積技術に関しては，映像の外部記憶装置への蓄積方法，再生時の音声と画像の同期問題，映像提示におけるリアルタイム性の維持など，システム的に興味深い問題があり，VOD技術と関連しているいろいろと研究が進んでいるが，ここでは議論しない．

2・2 映像メディアの検索と提示

検索は映像空間に含まれる情報をいかに取り出すかという問題である．検索手法は，図4に示すように，この空間の任意の断面で表現できる．当然，クエリーもマルチメディアデータを利用して行う．映像に対しては以下のような検索手法が考えられる．

テキストによる検索

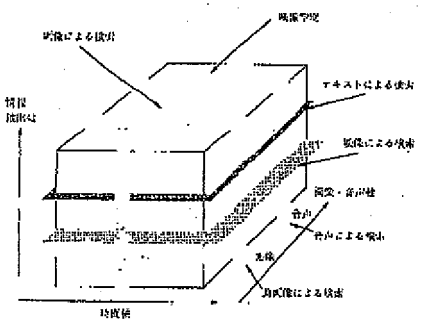

図4 検索の手法

画像による検索
音声による検索
映像による検索(動画像だけも含む)

これらの検索手法は排他的ではなく，共存が可能である．例えば，男女がペアで歩いている映像を検索する場合のクエリーとして，

クエリー1：テキスト：「男と女が一緒に歩いている」
クエリー2：画像：男女がペアの画像
クエリー3：動画像：男女がペアで歩いている動画像

などが考えられる．これらを同時に与えれば，システム側は有効な情報が多くなるので，良い検索結果を提示できるはずである．このように考えると，クエリーはデータのフィルタの役目を果たしている．これは画像データベース場合とまったく同様である[Faloutsos 94, Mukunoki 94a]．

クエリーをどういう形でシステムに与えるかは問題である．テキストで与える場合は，マルチメディアオブジェクトとその関係を言葉で記述する．当然，人間の動作や音声の特徴なども記述可能である．対象物の動きや色，形などをもとに検索する場合は，言葉で表現するよりも，データそのものをクエリーとして利用するほうが便利である．この場合のクエリー処理は，与えられたメディアデータに対して「似たもの」を探す処理となる．一般的には，データにはさまざまな解釈が可能であるので，画像，音声，動画像などが似ていることをどのように定義するかは，大問題である．

次の問題は，検索された映像データの提示方法である．映像情報は時間軸を持っているので，その内容を確認するためには，それなりの時間がかかる．検索されたものが目的に合っているかどうかを判断するためにはブラウジング機能が必須になる．これは，どういう形でクエリーが発せられたかにも依存する．また，検索がうまくいかなかった場合，システムがどのように推論したのかをユーザに示すことも重要である[前田 95]．

2・3 知的映像検索

「知的な検索」とは何を意味するのであろうか？ クエリーが曖昧でも適当に検索してくれるような検索を想像する人もいれば，ある種の推論を用にはさんで検索してくれるようなシステムを想像する人もいると思われる．本稿では，データベースへのデータの登録時にそれほど手間をかけなくても，それなりに検索できるようなメカニズムをさして「知的な検索」と考える．

この意味での知的な検索ができるシステムを実現するうえでの技術的な課題は次の3点である．

(1) 信号として与えられるマルチメディアデータから検索手法を想定したさまざまな検索のためのインデックスを作成する技術．言い換えるならば，映像空間の情報抽出次元に対するデータを自動生成する技術で，これはパターン認識の分野の技術である．この処理は，認識誤りが本質的で完全な認識は期待できないことを意識しなければならない．

(2) インデックスが認識誤りを含んでいても，重大な検索ミスを起こさないような検索手法が必要である．完全なデータベースから不完全データベースへと進んできたクエリー処理を，誤りを含むデータベースに拡張することが必要である．

(3) データ間の似ている度合い(類似度)の定義がユーザによって異なるので，個人に適応できるように，検索の過程において類似度計算が定義できるメカニズムが必要である．また，システムが行っている推論の内容をユーザに提示する手法も，知的検索システムには不可欠である．

現在報告されている映像処理，検索システムでは，これらの点はそれほど意識されてはいないが，今後は重要になると考える．

3. 映像の構造化技術

一連の映像は，図5に示すように，1枚の静止画像とは違って，作成者が意図したストーリーを持っている．このストーリーは，いくつかのサブストーリーに分割でき，さらにサブストーリーも分割できる．このように映像は，一般的には階層構造をしており，ツリー構造データで記述できる[柴田 95]．

このツリーの葉にあたる部分は一連の意味を持つ連続したフレームの系列であり，ショットと呼ばれる．ショットの切れ目をカットと呼ぶ．これらの単語は同じことを違った側面から表している．

映像を処理・蓄積して知的検索を実現するためには，基本的には，映像をショットに分割する必要がある．映像を計算機で扱う利点は，VODなどのように同時アクセスに対するサービスの多重化や同じ映像の異なる部分の再生など，従来の記憶媒体に比較していろいろあるが，やはり，意味的な処理を実現するためには意味のある最小単位であるショットを抽出する処理は不可欠である．

映像の構造化処理としては，ショットを抽出した後に，それらを分類し，インデックスを付加しておく必要がある．そのためには，ショットから特徴量を抽出したり，ショットを分類する処理が必要となる．

基本的には，情報抽出箱のデータを作成する処理は，画像を対象としたデータベースで行われてきたものをショットに対して適用する処理になる．ただ，映像は同期した音声や時間軸を持っており，次元数が多くなるので，データ量も多く，計算機で扱うのが画像ほど簡単ではない．これに伴って，検索方法も多様化する．また，結果を提示する方法も映像特有のものが必要であり，ヒューマンインタフェースに重点を置いた研究が活発に行われている．

3・1 映像からのカット検出法

人間が映像を見た場合，カットはかなり正確に検出できる．この操作において，人間は意味的な情報と信号的な情報を統合的に利用していると考えられる．

計算機にカット検出を行わせる場合は，信号的な情報のみに頼らなければならない．そのためには，カットとは何かを画像の信号レベルの言葉に置き換えるカットのモデルを構築する必要がある．もちろん，逆に，ショットのモデルを考えても差し支えない．

カットのモデルとしては以下のようなものがある[大辻 92]．

(1) カットの前後で画素値の分布が時間的に不連続に変化する．
(2) きわめて短い時間に大きな変動量が発生し，直後に収束する．
(3) 変動は画面の広い範囲に及ぶ．

多くのカット検出法は，これらのモデルの一つまたは複数のモデルを利用している．これらのモデルをどう実現するかでさまざまなカット検出法が提案されている．例えば，(1)のモデルを利用する場合でも，画面の輝度ヒストグラムの差分を用いる方法[外村 89, 長坂 92b]もあれば，フレーム間差分に基づく方法[大辻 91]も考えられる．しかし，本質的に同じモデルに基づくので，手法的にそれほど大きな違いはない．これらの手法の比較検討については[大辻 92]を参照されたい．

これらの手法は，さまざまなパラメータを持っており，その調整によりカット誤検出率，未検出率が変化する．[大辻 92]によれば，カット誤検出率＝カット未検出率となるようにパラメータを調整すればどの手法でもカット誤検出率を3%以下に抑えられる．

ここであげたカットのモデルは，フェードアウトや特殊効果による徐々に変化するカットの検出には失敗する．これに対処するために，カットの前後数フレームでの画像データの統計を利用した手法[Zhang 93]が報告されている．

たとえどのようなモデルを作成しても，セマンティックギャップの問題があり，完全なモデルの生成は困難である．意味的な情報に基づくカットが存在する限り，信号的な手法では検出はできない．したがって，これらのカット検出法は，誤りを本質的に含むものとして，扱うことを考えていかなければならない．

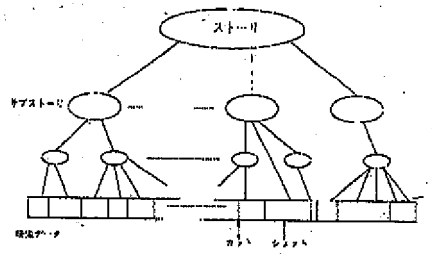

図5 映像データの階層

3・2 ショットからの特徴抽出

何らかの形でショットが抽出できると，次の課題は抽出されたショットから検索や編集処理に役立つ特徴量を抽出することである．特徴量として現在までに利用されているのは，撮影時のカメラの動きに関する情報，撮影者や記者がつけたアノテーション，ショットに含まれる色情報などである．

撮影時のカメラの動きは，オプティカルフローに基づいて簡単に検出できる[上田 92]．これはカメラの動きに対応して出現するオプティカルフローのパターンが決まっていることを利用している．ただし，オプティカルフローの計算は誤差が大きく，それほど信頼性は高くない．

[Tonomura 94]では，時空間スペースを用いてカメラの動きを検出している．これも，カメラの動きが時空間スペースである種のパタ・ンになることを利用している．

カメラの動きの特徴は，簡単に抽出できるが，この情報がそのままの形で検索に有用な情報を提供するかどうかは疑問である．これに対して，映像に入っている文字情報を抽出しようとする研究も試みられている[熊谷 95，滝沢 95]．映像は人間がつくるのであり，重要な情報は文字の形で映像に組み込まれているので，うまく抽出ができれば，特徴としては有効なものになる．

[柴田 95]では，放送用の映像を対象に記者がつけたアノテーションを人手で入力し，特徴量とすることを提案している．これは，制作者がつける情報であり，検索のためには大変有用なものになる．また，これらの情報がカットの自動検出にも応用できることを示しており，意味的なカット検出の手法としても注目できる．

[Yamane 93]では，スポーツ映像を区別するために色情報をショットから抽出している．

これらの特徴量は単にコンテントを表すだけでなく，他のショットへのリンク情報になることもある．例えば，同じ色の同じ配置が，異なるショットのフレームで検出された場合，これらのショットには同じ物体が現れている可能性が高い[長坂 92a]．このような場合は，この情報をショットからショットへのリンク情報として利用する．

ショットから特徴量が抽出されると，それらを似たものどうし集めることにより，クラスタリングを行う．これは検索時のインデックスとしても利用できる．特徴空間の類似度をどのように定義するかは大問題であるが，便宜的に，特徴空間をユークリッド空間と考えて，そのなかで距離を定義して類似度を計算する[柴田 95]．ショットのクラスタリングにおいては，時間軸の制約があるので，時間的に隣接するショットのみがクラスタリングの対象となる点にも注意を要する．まだ，多くの映像を蓄積したデータベースは存在しないので，クラスタリングに対してはそれほど研究が進んでいないが，今後重要になることは間違いない．

4. 知的映像検索技術

現在のところ，大規模に映像データを蓄積してそれを検索しようという試みはほとんどなく，実験的に検索の可能性を調べているのが現状である．映像の検索

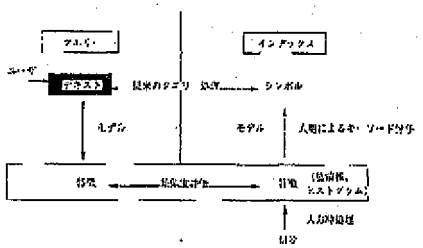

図6 テキストによる知的検索

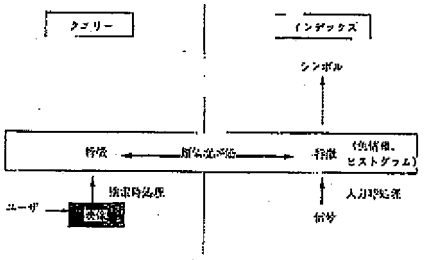

図7 映像による知的検索

方法は大きく分けて2種類考えられる．図6に示すテキストによる検索と図7に示すパターンによる検索である．

テキストによる検索では，インデックス側に同じレベルの記述がモデルを介して自動的に，もしくは，人間により付与されている場合は，従来のデータベースの検索法と同じ枠組みになる．付与されていない場合は，与えられたクエリーによるテキストから，何らかのモデルを介して特徴レベルの記述を生成し，類似度による評価を行う(図6参照)．

[Yamane 93]ではスポーツのシーンを色の組合せで表すモデルを作成し，それによってサッカーや相撲の映像を検索するシステムを作成している．このシステムでは，データベースの映像は色で識別できるものをうまく選んでおり，応用はかなり限定される．しかし，うまく状況を設定して初期的な知的検索を実現していると例と言える．

映像をテキストで検索しようとすると入力時に画像認識の問題に直面する．画像認識は，困難な問題であり，現在の技術では，さまざまな状況に対応することはできない．これに対して，画像から特徴量を抽出する処理はかなり汎用的なので，このレベルの記述でマッチングをとることができれば，汎用的な知的検索が実現できる可能性がある．

映像や画像そのものをクエリーとして利用する場合は，図7に示すように，検索時に入力時に行ったのと同じ方法で特徴を抽出する必要がある．この場合も，最終的には特徴レベルでの類似度評価に帰着する．

映像は，画像と違って，時系列のデータを含んでおり，音声も存在する．これらのなかから映像を識別するうえでの特徴を探す試みが行われている．[岡本 95]では，人物の映像を検索するのに，クエリーとして映像を与えるという枠組みで，キーフレームの選択方法，静止画特徴と動き特徴の統合について検討している．また，[高橋 94]では，ジェスチャスポッティングのために，動画像を用いた動画像のマッチング方法について検討している．

音声を本格的に扱った研究はまだ少なく，厳密な意味では，まだ，映像の処理は行われていない．今後，音声処理を統合した本当の映像処理の研究が出てくることを期待したい[Smoliar 94]．

映像空間は情報提示物を除いても4次元の空間であるので，検索技術にとって，その情報の提示技術は重要である．映像メディアとの対話機能を重要視して，映像構造をユーザにわかりやすいように見せることにより，人間が検索を行ううえでの道具と位置づける立場から研究が進められている．

一連の映像データを対象に，そのなかのショットがどのような構造になっているかを視覚的に表示することにより，検索を支援しようとするシステムがいろいろと提案されている[Smoliar 94，Tonomura 94，上田 93]．これらは，ショットを3次元のiconと呼ばれるアイコンで表示する．次元は画像軸2次元と時間軸1次元である．時間軸に対する立体の大きさで継続時間がわかり，画像軸に対応する面にそのショットを代表するキーフレームを表示する．その他，画像を時空間表示してその断面を見せたり，キーフレームだけを2次元に並べてその内容を一覧できるようにするなどさまざまな工夫がなされている．これらの技術は，ヒューマンインタフェースの観点から議論されており，検索結果として，候補映像が得られた場合には強力なツールになる．

[長坂 92a]では，画面をブロックに分割し，各ブロックの色情報とその配置に基づき，同じ色が同じ配置で現れるフレーム間にリンクを張ることを提案している．これは，このようなフレームは同じオブジェクトを含んでいる可能性が高いという仮定に基づいている．同じオブジェクトが現れたフレームを色で区別することにより，ショット間の関連が一目でわかるインタフェースを作成している．この手法が有効であるかどうかは映像の内容にもよるが，うまく使えばそれなりの効果は上げられる．

5. おわりに

本稿では知的映像メディア検索技術と，その検索を想定した入力系の技術を中心に解説した．パソコンがようやく映像を表示できるようになり，映像の処理が現実的なものとなってきた．現在行われている研究は，原理を追求するというより，現状のマシンで，実際に映像を扱って，その枠内でできる簡単な処理で，できることだけを行おうとしている感がある．これに対して，人間を支援する立場からのインタフェース，コミュニケーションに目的を置いた研究は着実なステップである．

映像の処理，蓄積，検索，提示の技術がマルチメディアの中心的存在になることは明らかであり，さまざまな応用が考えられる興味深い分野である．知的映像メディアの検索技術は人間にとっては夢であり，バラ色の世界である．しかし，この裏には，パターン認識・人工知能の壁の問題がひそんでいる．認識技術，人工知能の技術は完全でないことを十分意識したうえで，検索技術，提示技術をうまく活用していくことが，この分野を発展させるキーポイントであると考えている．

今後，映像に関する知的処理，知的検索技術がますます発展することを期待する．

◇参考文献◇

[Davenport 91] Davenport, G., Smith, T. A. and Pincever, N.: Cinematic Primitives for Multimedia. *IEEE Computer Graphics & Applications*, Vol. 11, No. 4, pp. 67-75 (1991).

[Faloutsos 94] Faloutsos, C., Flickner, M., Niblack, N., Petkovic, D., Equitz, W. and Barber, R.: Efficient and Effective Querying by Image Content. *J. Intell. Inf. Syst.*, Vol. 3, No. 4, pp. 231-262 (1994).

[Grosky 94] Grosky, W. I.: Multimedia Information Systems, *IEEE Multimedia*, Vol. 1, No. 1, p. 12-24 (1994).

[植田 94] 植田，美濃：知能情報メディアへの期待，1994信学会秋季大会，pp. 483-484 (1994).

[熊谷 95] 熊谷，仙田，美濃，池田：手話の功能的・空間的特徴を利用した動画像からの手話領域の抽出，1995画像電子学会

年次大会, pp. 23-24 (1995).
[Mackay 89] Mackay, W. E. and Davenport, G.: Virtual Video Editing in Interactive Multimedia Applications, *Commun. ACM*, Vol. 32, No. 7, pp. 802-810 (1989).
[前田 95] 前田, 榎木, 美濃, 池田：検索者の意図を的確にくみとる画像検索システム, 1995 電子情報通信学会秋季大会, A-173, pp. 176 (1995).
[Mills 92] Mills, M., Cohen, J. and Wong, Y. Y.: A Magnifier Tool for Video Data, *Proc. CHI'92*, pp. 93-98 (1992).
[Mukunoki 94a] Mukunoki, M., Minoh, M. and Ikeda, K.: Retrieval of Images Using Pixel Based Object Models, *Proc. 5th IPMU*, Vol. 2, pp. 1127-1132 (1994).
[榎木 94b] 榎木, 美濃, 池田：画像認識結果をインデックスとする風景画像検索, 画像電子学会研究会, No. 141, p. 1-4 (1994).
[Nagasaka 92] Nagasaka, A. and Tanaka, Y.: Automatic Video Indexing and Full-Video Search for Object Appearances, *Visual Database Systems*, Vol. II, pp. 113-127, Elsevier Science Publishers (1992).
[長坂 92a] 長坂, 宮武, 上田：動画像情報リンク付けの一下法, テレビ技報, Vol. 16, No. 10, pp. 25-30 (1992).
[長坂 92b] 長坂, 田中：カラー・ビデオ映像における自動索引付け法と物体探索法, 電通学誌, Vol. 33, No. 4, pp. 543-550 (1992).
[岡本 95] 岡本, 美濃, 池田：人物動画像の検索における動画像キーの利用と履歴特徴の統合, 信学技報, IE94-131, PRU94-131, p. 9-16 (1995).
[大計 91] 大辻, 外村, 大原：速度情報を使った動画ブラウジング, 信学技報, IE90-103 (1991).
[大計 92] 大辻, 外村：映像カット自動検出方式の検討, テレビ技報, Vol. 16, No. 43, pp. 7-12 (1992).
[Pentland 94] Pentland, A., Picard, R., Davenport, G. and Haase, K.: Video and Image Semantics: Advanced Tools for Telecommunications, *IEEE Multimedia*, Vol. 1, No. 2, p. 73-75 (1994).
[坂内 92] 坂内三夫：画像・マルチメディアデータベース, テレビジョン学会誌, Vol. 46, No. 11, pp. 1474-1479 (1992).
[Sakauchi 94] Sakauchi, M.: Database Vision and Image Retrieval, *IEEE Multimeia*, Vol. 1, No. 1, p. 79-81 (1994).
[柴田 95] 柴田正啓：映像の内容記述モデルとその映像構造化への応用, 信学論, Vol. 78-D-II, No. 5, pp. 754-764 (1995).
[Smoliar 94] Smoliar, S. W. and Zhang, H.: Content-Based Video Indexing and Retrieval, *IEEE Multimedia*, Vol. 1, No. 2, p. 62-72 (1994).
[高橋 94] 高橋, 関, 小島, 河：ジェスチャー動画像のスポッティング認識, 信学論, Vol. 77-D-II, No. 8, pp. 1552-1561 (1994).
[滝沢 95] 滝沢, 笹田, 美濃, 池田：動画像からの看板文字パターン列の補間, 信学技報, IE94-133, PRU94-133, pp. 25-32 (1995).
[田村 95] 田村, 池田：物理的制約を導入した仮想空間でのシミュレーション, 知能情報メディア, pp. 105-126, 総研出版 (1995).
[外村 89] 外村, 安部：動画像データベースハンドリングに関する検討, 信学技報, IE89-33 (1989).
[Tonomura 94] Tonomura, Y., Akutsu, A., Taniguchi, Y. Suzuki, G.: Structured Video Computing, *IEEE Multimedia*, Vol. 1, No. 3, pp. 34-43 (1994).
[Ueda 91] Ueda, H., Miyatake, T. and Yoshizawa, S.: IMPACT: An Interactive Natural-motion picture Dedicated Multimedia Authoring System, *Proc. CHI'91*, pp. 343-350 (1991).
[上田 92] 上田, 宮武, 吉沢：認識技術を活用した対話型映像編集方式の提案, 信学論, Vol. 75-D-II, No. 2, pp. 216-225 (1992).
[上田 93] 上田, 宮武, 長野, 長坂：動画像解析に基づくビデオ構造の視覚化とその応用, 信学論, Vol. 76-D-II, No. 8, pp. 1572-1580 (1993).
[Yamane 93] Yamane, J. and Sakauchi, M.: A Construction of a New Image Database System which Realized Fully Automated Image Keyword Extraction, *IEICE Trans. Inf. & Syst.*, Vol. E76-D, No. 10, pp. 1216-1223 (1993).
[Zhang 93] Zhang, H., Kankanhalli, A., Smoliar, S. W.: Automatic Partitioning of Full-motion Video, *Multimedia Systems*, Vol. 1, No. 1, pp. 10-28 (1993).

著者紹介

美濃 導彦

1978年京都大学工学部情報工学科卒業, 1980年同大学院修士課程修了, 同年, 京都大学工学部助手, 1989～95年マサチューセッツ大学アムハースト校客員研究員, 1995年京都大学工学部情報工学教室助教授, 現在に至る. 博士(工学). 知的コミュニケーション環境の研究に従事. 工学博士, IEEE, ACM, 電気学会, 電子情報通信学会, 日本ロボット学会各会員.

論　文 (Technical Papers)

人工知能学会誌　第11巻 第1号 (通巻65号) 1996年(平成8年)1月1日発行　(c) 人工知能学会 1996
編集兼発行人 〒162 東京都新宿区津久戸町4-7 OSビル402号室 (社)人工知能学会 佐々木三千矢 Tel.03-5261-3401 Fax.03-5261-3402 (振替 00150-6-103586)
印刷 〒101 東京都千代田区神田錦町3-1 株式会社オーム社 (振替 00150-6-20018)　製版印刷協力 株式会社エヌ・ピー・エス
ISSN 0912-8085　発売元 オーム社　定価 2100円(本体 2330円・税 70円)